BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

SECTION VI: TOPICS IN DRUG DESIGN AND DISCOVERY

Editor: Michael C. Venuti
Parnassus Pharmaceuticals, Inc.
Alameda, CA 94501

Chapter 29. 3D Database Searching and Docking Strategies

Christine Humblet and James B. Dunbar, Jr.
Parke-Davis Pharmaceutical Research
Division of Warner-Lambert Company
Ann Arbor, Michigan 48105

Introduction - Traditionally, chemical information systems have relied upon two-dimensional (2D) connectivity tables as a means to represent chemical structures. Graph theory (*i.e.*, subgraph isomorphism algorithms, clique detection methods), sometimes coupled with distance geometry approaches (*e.g.*, difference distance matrix), have and continue to be used for structure, substructure and similarity searching driven by geometric or property-based queries (1-5).

Recent advances in computer technologies and dramatic cost reductions for physical memory and mass storage devices now allow for extensive data storage. Simultaneously, modern workstations have undergone a major expansion in computational power, providing the capability to access the hundreds of thousands of data records needed to search and analyze structural as well as conformational information. As a result, impressive progress in the development of automated methods for generating three-dimensional (3D) structures from traditional 2D chemical diagrams has occurred.

The advent of straightforward 2D to 3D structural database conversions has prompted the simultaneous development of methods to retrieve, classify (or cluster), and analyze structural information, especially as they relate to the interrelationships of structures to biological activities. Recent biological technologies have further emphasized the fundamental value of molecular structures, constitutive building blocks that can clarify drug-receptor associations and help understand biological functions of receptor proteins. As reviewed in this chapter, structural databases can provide invaluable resources to medicinal chemists, offering an expanding range of tools to enhance the drug discovery process, including methodologies in information retrieval, knowledge-based strategies, multiple docking procedures, and *de novo* design approaches.

STRUCTURAL DATABASES AND SIMILARITY SEARCHES

Similarity searches initially involved geometric matching of structural groups, a simple strategy using query patterns based on the atom connectivity of the 2D chemical structures stored in databases. For example, searches are commonly used to verify that a particular chemical structure is indeed novel (*e.g.*, patent or spectral search)(6). Common database searching strategies have been reviewed in the recent literature (1,4,7).

2D Database Searches - Such database searches are global in nature, since the query is meant to identify structures that present total or partial similarity. When dealing with structures that relate to biological activities, however, the database queries need to incorporate geometric features to reflect on the presence of particular structural elements that appear to impact on the biological activity, *i.e.*, the pharmacophore or pharmacophoric moieties. A pharmacophore can thus consist of non-connected atoms spread throughout an active structure. It can also be an ensemble of atom types and their distance relationships, pharmacophore moieties (*e.g.*, presence of a particular heterocyclic moiety, H-bond acceptor or donor, positive or negative charges), or receptor site points. Database searching must therefore provide the flexibility to identify crucial substructural or geometric

elements. The methods of subgraph isomorphism and distance geometry have attempted to address these needs as amply reviewed in the recent literature (4,5,8).

The majority of the 2D searching methods retrieve only exact query matches, overlooking potential hits (9). Newer methods are currently being developed to overcome this limitation. For example, a two-step approach involving unique subsimilarity screening followed by an approximate maximal common subgraph matching step was shown to find substructures that nearly match (*i.e.*, "fuzzy match") the initial query (9).

2D to 3D Database Conversion - The program MOLPAT was among the first to introduce 3D concepts in database searching (10). Evolving from 2D substructure search systems, the first generation of 3D similarity searching systems were aimed at a structural geometry similarity search based on an input target structure. The database search procedure involved a two-level retrieval algorithm in which the time-consuming geometric search was preceeded by a rapid screening search based on interatomic distance ranges (4,5,8). The atom mapping technique is an example where the degree of resemblance between pairs of 3D structures is calculated from their interatomic distance matrices (8,11,12). The computational efficiency of the method was recently assessed and compared to others (8,11).

The emergence of powerful molecular graphic systems has intensified the desire to expand database retrieval mechanisms beyond 2D molecular connectivities to incorporate 3D properties in the queries. Today, most database retrieval systems have been upgraded to 3D structures as examplified in the recent availability of software tools from various commercial sources such as Cambridge Structural Database (CSD)(13) with the QUEST3D and VISTA programs (14), Chemical Design Ltd with MACCS3D (6), Molecular Design Limited with CHEMDBS3D (15), Tripos Associates Ltd. with UNITY (16), Chemical Abstract System (17), and Fine Chemical Directory (FCD) (6).

The rule-based program, CONCORD (18) is the most popular procedure to convert existing 2D databases (19). Using artificial intelligence methods, an alternative approach has been developed in COBRA (20), a successor to WIZARD. This program decomposes structures into "simple conformational units" and progressively assembles reasonable "subconformations" for each unit. The conformational space of a molecule is then represented as a tree-like structural assembly which lends itself to graph-searching techniques. Like all other knowledge-based systems, this method is limited by the initial dictionary of conformational units considered. Efforts to incorporate distance geometry template generation to address this limitation are anticipated.

3D Conformational Flexibility - An important issue in 3D database system development rests with the ability to rapidly and efficiently address conformational flexibility for 3D structures converted from a 2D connectivity table. At this point, there is no universally accepted solution to this problem. For example, conformational information may be stored explicitly in the database in the form of multiple conformations or implicitly in the form of interatomic distance ranges. In the explicit approach, it becomes necessary to preload multiple conformations into a database. Major difficulties lie in the selection and number of conformations to retain for each structure. Not only does this impact the size of the database and the speed of the search, but also the successful identification of relevant conformations (7). In the implicit approach, the conformational flexibility is inferred by multiple interatomic distance ranges, involving distance geometry algorithms, and is widely illustrated in programs such as ALADDIN (21) or EMBED (22). The use of bounded distance matrices, a combination of distance geometry and graph theory methods, was shown to be an efficient means to address the conformational flexibility while matching pharmacophoric patterns in a 3D database (23). Conformational flexibility may also be taken into account while the database search is being performed. Flexible search queries or "on-the-fly" conformational analyses have been reported (24). The Sybyl UNITY database system illustrates another recent initiative to incorporate newer methods such as a tweak algorithm that allow for conformational expansion (16).

STRATEGIES IN 3D SIMILARITY SEARCHING

In the absence of a receptor structure, 3D databases can be searched indirectly, matching the known 3D structures of various ligands on the basis of 3D similarity. This computerized screening is used to identify structural leads (25).

Grid-based Similarity - As originally introduced by Carbo, molecular similarity between two molecules can be determined from the electron density derived from quantum mechanical calculations (26, 27). The technique has since been extended to use other structural properties

such as electrostatic potentials, electric fields, and shape. In the program ASP, for example, molecules are surrounded by a rectilinear grid; the structural property is evaluated at each grid intersection and integrals are evaluated numerically or by utilizing Gaussian functions rather than grids (28-32). The similarity index resulting from these calculations provides a means to quantitatively relate the molecular similarity to the observed biological activities. In the program MEPCONF, molecular similarity is derived from the molecular electrostatic potential (MEP) mapped on a grid surface that is proportional to the van der Waals surfaces of two molecules (33). Both rigid rotations and translations of one system relative to another and internal conformational expansion of the relative molecules are carried out in an iterative procedure to maximize the similarity. The similarity is measured using the Spearman rank correlation coefficient on pairs of points on the common, coincident, grid computed at each iteration.

Surface and shape similarity - The 3D similarity can also be derived from molecular surfaces or shape analyses. For example, a 3D shape definition has been described by considering points on spheres that surround the molecules (3,36). In this gnomonic projection, 3D molecular properties such as size and electrostatic potentials can be assigned as scalar or vector functions.

In a distinct approach, SPERM, an icosahedral mapping and matching procedure (37) has been used to specify molecular properties surrounding molecules (38, 39). 3D molecular shape is the primary matching parameter and requires an initial superposition of centroids or center of mass for the molecules compared. A 3D space scanning is applied to the molecule being compared to a reference template, leading to a shape similarity score evaluated from the sum of squares of the differences (SSD). Any chosen molecular property such as molecular electrostatic potentials or electron density can then be further quantified using either Carbo (26, 27) or Hodgkin (29) similarity indices.

Another method has been derived from the use of approximate wavefunctions (40). The shape of a molecule is defined by exploring the molecular surface with a helium atom probe. The calculation involves the evaluation of a simple expression for the repulsive interaction between molecule and probe. The similarity index is calculated from the overlap integrals between the orbitals of one molecule and those of the other.

Following an initial atom-matching alignment, the molecular shapes have been used to optimize the alignment (41). A surface comparison is evaluated from all surface dots within a certain distance of the corresponding dots of the mean surface. It has been used in an effort to compare biologically active or inactive compounds belonging to a similar structural type.

Two hashing schemes, one based on connectivity and the other based on internal distances, have been applied to the problem of similarity and clustering of shape data (42).

Shape similarities have traditionally used geometrical representations, but topological representations have also been explored (43). The shape group method has been extended to utilize density domains of molecules analyzed at different electron density values (44). The electron density is calculated by an *ab initio* method. The shape codes generated in the analysis can then be used to provide a similarity measure for two molecules in a topological rather than a geometrical description.

Pharmacophore-based Similarity - In the majority of the 3D similarity studies, a preliminary molecular alignment is necessary. Usually, geometric features derived from pharmacophoric hypotheses have been used to guide manual superimpositions of corresponding molecules, sometimes handled through molecular graphics manipulations. Recent studies illustrate the trend to automate molecular superpositions, building upon the 3D database retrieval techniques to limit biases often introduced by individual users.

In the program SUPER, the correspondence of molecules is achieved through optimum superposition of two molecules using their overlap integrals (45). SUPER is designed to identify the best twenty superpositions by generating a network of grid points at the van der Waals surfaces of the two molecules and assigning united atom potentials to these points. The search for the superposition is carried out by rigid body translations and rotations such that each atom of molecule 1 is eventually paired with each atom in molecule 2. SUPER discards matches of points if the difference between two potentials is above a certain set value. This is also used as a measure of the goodness of fit.

The RECEPS program, extended with a module AUTOFIT, is designed to superimpose two molecules on the basis of hydrogen-bonding patterns (46). This is accomplished by an iterative least squares superposition procedure with minimization of both position and direction vectors, checking that the appropriate hydrogen-bonding heteroatom types are correctly matched. This method is capable of conformational expansion based on systematic rotation of torsion angles, and is therefore subject to the combinatorial explosion problem.

In another approach, ligand structures are used to define potential receptor binding sites, the anchor points (47). Algorithms are then applied to identify binding partners for these anchor points and to position them in 3D space. The ensemble of binding partners constitutes a "pseudo-receptor" model for the ligands compared. YAK, a receptor mapping program, allows the orientation of binding partners according to molecular interactions. This program is used in combination with a database of 110 atom types defined on the basis of their interaction properties.

DISCO, based on a clique detection method, determines the superposition of a series of molecules and proposes pharmacophore hypotheses (48). The location of ligand and receptor site points required in the superposition rule, as well as the selection of multiple conformer sets, are prepared with the program ALADDIN (21). DISCO can search for superposition of a set of molecules, or by relaxing tolerances for a match, providing insight into alternate modes of superpositions.

In an approach that can be extended to receptor mapping, hydropathy has been used to measure similarity based on complementarity (49). Hydropathic character may be examined by generating maps for the ligand (KEY program) or for the receptor (LOCK program). The program LOCKSMITH utilizes the maps generated by the KEY program and searches for orientations and common structural features for a series of molecules.

INTERMOLECULAR DOCKING STUDIES

The availability of 3D structural databases has driven the development of methods to locate and analyze protein cavities. More recently, new algorithms have been described to improve the simulation and evaluation of ligand-binding site intermolecular associations as well as protein-protein interactions.

Binding-site Surface Depiction - In CAVITY, a set of algorithms allows the user to isolate and define the extent of a particular binding cavity (50). Solid modeling techniques are employed to produce a detailed cast of the active site region. This method necessitates the selection of an initial seed point in the binding site. Electrostatic and steric properties can be represented on the cavity surface to assist the graphical analysis of the detailed molecular interactions between ligand and receptor surface.

Other approaches to depict the binding site boundaries use different algorithms and present varying levels of accuracy and computer requirements, but all allow automatic identification and visualization of potential binding sites in proteins (51-53). Once the potential binding sites have been located, other programs can be used to examine ligand docking.

Graph Theoretical Techniques - To determine whether any of a given series of compounds are able, on geometrical grounds, to interact with an active site of interest, Smellie and coworkers (54) applied a bipartite graph representation that takes advantage of the clique detection algorithm of Kuhl (55) associated with a clique filtering method. This docking strategy necessitates the identification of matching sites and can be applied to pharmacophore or binding site queries. Conformational flexibility can be incorporated in the form of upper and lower bound distances and distance constraints can be imposed during the docking procedure. This method holds promise as an efficient first pass tool to prepare large 3D database searches. Also based on graph theoretical techniques, a method was described to convert the surfaces of the two (macro)molecules into graphs (56). Appropriate parts of the two surfaces are then matched on the basis of corresponding graphs or patterns.

Shape Complementarity Methods - The program DOCK represents an early attempt to match "keys" into a known "lock" (57,58). 3D databases are searched for potential ligands that are complementary in 3D shape to a known receptor binding site and scored based on their goodness of fit. DOCK has been applied to a series of 103 inhibitors of chymotrypsin and correctly identified the two best known inhibitor structures (59). The authors also reported the screening of the 57,128 compounds available in the FCD; out of the 23 high DOCK-scored ligands tested, two structures

were found to inhibit chymotrypsin. In a more recent version of the program, DOCK2, better sampling and a more systematic searching of the orientation space have been reported (60). DOCK2 also includes a lattice-based method for evaluating the goodness of fit. These new features have been assessed in a variety of applications. DOCK was recently applied to the crystal structure of *L. casei* thymidylate synthase (TS) using the FCD (61). The TS substrate deoxyuridine monophosphate (dUMP), and several known nucleotide and non-nucleotide inhibitors of the enzyme, were identified and adequately scored by the computational screen. Additional molecules unknown to bind TS (*e.g.*, sulisobenzone), were also retrieved from the database and shown to inhibit the TS enzyme.

An approach using bidimensional surface profile to find molecular shape complementarity reduces the complex 3D surface to a 2D representation called "angular profiles". The observed crystal-bound conformations of kallikreine and bovine pancreatic trypsin inhibitors were reproduced, but it was not the best matching solution, suggesting that this method could be used as an initial step toward a 3D study of the shape complementarity between two molecular surfaces (62).

An alternative approach, patterns of points or webs are generated to represent the two molecular surfaces (63). McLachlan's least squares fitting method (64) is then used to match corresponding webs in the two surfaces. Local complementarity and van der Waals interaction energy are used as filtering criteria to select multiple docking orientations. These are then scored based on electrostatic interaction energy. The authors suggest that this method is amenable to coarse dihedral sampling to address conformational flexibility of the ligand.

Another method that uses similarity in the design of compounds involves determining the electrostatic properties of the ligand in the field of the receptor. If electrostatic characteristics are important, then their determination will allow searches for molecules that have similar charge characteristics. A program, YING, determines the point charges of a ligand bound to a receptor based on maximizing the complementarity of the charges of the ligand and receptor (65). The method uses the van der Waals surface of the ligand, the electrostatic potential of the receptor at that surface, and the partial charges of the ligand atoms. The program then maximizes the charge complementarity while still maintaining any formal charge associated with the ligand. Ligand and receptor atoms are kept fixed during this process.

Metropolis Monte Carlo Method - In the program DISDOCK, a simulated annealing procedure is applied to position and orient the interacting molecules, which are both treated as rigid bodies (66). The distance constraints that are necessary to guide the relative intermolecular orientation are derived from a selection of atom pairs such as a hydrogen donor and its complementary acceptor atom. When applied to a set of serine proteinase complexes, the best results are obtained when known distance constraints are available. The results also point to a dependence on the starting ligand conformation.

In AUTODOCK, a Monte Carlo conformational analysis is combined with rapid energy evaluation using molecular affinity potentials (67). A limited number of rotatable bonds can be randomly varied in the ligand structure. Although this approach does not require preliminary knowledge of the binding site, it appears that the substrate binding modes can be affected by the starting orientation. In an alternative strategy, rigid molecular fragments are docked into the binding site (68). The unique feature of this method is the ability of buried fragments to float to the binding pockets from inside a grid representation of the receptor as opposed to placing the ligand at some arbitrary position well outside of the receptor grid. This is accomplished by a scoring function that measures the average distance to the surface points on the grid. Simulated annealing is then used to complete the simulation.

Another method utilizes a two step procedure. A predocked ligand is submitted to an initial coarse sampling (*i.e.*, rigid body translation and rotation) in positional space with limited conformational exploration. A Metropolis Monte Carlo (69) minimization is then applied with complete freedom of movement of the ligand to optimize the structures obtained from the first step. The receptor is rigid throughout the simulation and the method does not use a grid/surface based representation (70).

Protein-Protein Docking - Although the recent literature emphasizes docking procedures that involve a small molecule ligand binding to a macromolecule, molecular complementarity also applies to protein-protein recognition phenomena. Early attempts in protein-protein docking relied either on electrostatics or geometric description of the molecular surfaces (71).

The DOCK program has been extended to the area of rigid body protein-protein docking (72). In a similar approach, simplified protein models with one sphere per residue are subjected to simulated annealing using a crude energy function where the attractive component is proportional to the interface area (73). The procedure finds clusters of orientations in which a steric fit between the two protein components is achieved over a large contact surface.

In another approach, each protein is reduced to sets of surface and internal volume grid points (74). Molecules are then docked by matching surface grid points. Optimal orientations are those with maximum matching surface points and minimum overlapping volume cubes.

DE NOVO DESIGN

Ligand Structures - The program CAVEAT was among the first to utilize 3D database geometry features to retrieve novel structural templates (75). In a recent application, the CSD was searched to extract structures in which specified bond directionalities were maintained at the correct distances and angles to identify structural peptidomimetic templates. In a related approach, the FOUNDATION database system retrieves molecules based upon matching some minimum subset of elements of the query (76). The query elements include atom position and type, bonded and nonbonded atoms, RMS fit, subsets with an occupancy range, and volume restrictions. By using appropriate atom type definitions, general characteristics such as hydrogen bond donor/acceptor and hydrophobicity can be included in the query. The unique feature in FOUNDATION is the ability to search for any combination of some specified lower limit of the elements that comprise the query. This "fuzzy match" allows the retrieval of molecules that do not fit the entire query. This can be used as starting points in the design of novel ligands, or in the selection of compounds to be tested in an assay. Since molecular fragments can be in the input, searches for bridging elements to link the important elements of a pharmacophore hypothesis can be run. The resulting hit list can be ranked by number of matching atoms, RMS-fit, or steric complementarity if a volume constraint is used.

In the program GROW, peptidic fragments are placed and connected within the confines of the receptor active site (77). The template library contains amino acid fragment units in their multiple low energy conformations. From an initial seed point, all template fragments are evaluated for binding site fit. Scoring is based on a molecular mechanics potential, including a solvation term. The growth can take place without any restrictions or can be directed by the use of several control parameters.

LUDI is a rule-based system that makes use of contact patterns found in the CSD to perform automatic design of novel compounds (78). LUDI looks for hydrogen-bonding interactions with appropriate distance and angle characteristics, lipophilic-aliphatic and lipophilic-aromatic interactions. The fragments employed are small molecules progressively connected to fit a binding site. LUDI is now capable of starting from an existing fragment, allowing incorporation of features extracted from a known ligand. The scoring function takes into account the number and quality of hydrogen bonds and also the contact surface area.

Another method, DELEGATE (within the software package BUILDER), generates an irregular lattice of points within an active site (79). First a DOCK run is performed to place a number of appropriate molecules within the active site. This collection of molecules is used to construct the irregular lattice, fusing the atoms of the DOCK structures with correct distance and angles into a composite structure that fits the active site. Additional minimization is required to regularize the structure. Electrostatic, hydrogen bonding, and shape complementarity characteristics can be accomodated. This lattice can also be used manually to view the disconnected structures and prune/connect atoms to form a composite structure.

The database system CLIX (80) utilizes a series of GRID (81) runs to determine where various probes would have favorable interactions with a target receptor. This produces an ensemble of possible sites of interaction. The database is comprised of structures from the CSD. CLIX searches for at least three coincident favorable points of interaction between the grid and the ligand with appropriate steric fit. The fits are scored based on the sum of the interaction energies.

GenStar is designed to grow tetrahedral atoms into an active site (82). The input controls include the heavy atoms of interest at the active site, a seed point (either from the enzyme or from a known inhibitor), and maximum number of atoms required in the compound to be generated. A "closeness grid" is calculated around the seed point to check for steric contacts. A scoring scheme similar to that used in DOCK includes the capability to detect potential hydrogen bonding sites.

LEGEND builds a structure sequentially, starting from randomly selected atom types positioned with random torsion angles (83). Intermolecular interactions are neglected during the atom generation process. A candidate atom is selected automatically if it is not bumping either the enzyme or any previous atoms in the growing candidate molecule. After a structure is complete, charges are assigned to all atoms and the structure is energy minimized in the active site. From the many structures generated by LEGEND, a separate post-processing program, LORE, may be used to select the more interesting structures for graphical analysis.

In preparation for site-directed drug design, the CSD has been used to retrieve aliphatic fragments and their respective bond properties (84-86). A new algorithm to create diverse, irregular, and physically reasonable 3D linear atomic chains or molecular graphs has been described (87). These can be used to generate structural templates for joining up regions in an active site (87).

Protein Structures - The use of structural databases in similarity or commonality searching of protein sequences and structures has been an area of extensive research during the past decade. Applications include prediction of structure from sequence (homology- or knowledge-based), design or engineering of proteins, and 3D model building for ligand design. Recent research has focused on recognizing the structural information available in known structures or substructures. Knowledge-based systems have been developed to access, compile, and analyze protein structure properties such as residue features (*e.g.*, side-chain rotamers), secondary structure (*e.g.*, turn motifs), and tertiary folds (88-90).

The similarity searching methods commonly applied to chemical structures are now finding their first applications in the field of protein structures. For example, a subgraph isomorphism algorithm based on a clique detection procedure has been applied to locate structural patterns in pairs of proteins (91).

EMERGING TOPICS

Clustering Methods - The speed at which 2D structures are being converted into 3D structures has led to an explosion of structural data records based on available databases such as the FCD, CAS, or other proprietary databases. This has led to the need to select a subset of 3D structures to be considered in a database query. Cluster analysis or automatic classification is the name given to a range of techniques for the grouping of multidimensional datasets. Although differing in their individual details, these methods are based on the grouping together, or clustering, of the most similar objects or pairs of objects. The similarities are typically calculated using nearest-neighbor searching routines (92). Various similarity searching strategies have been applied to identify an adequate subset of large databases for biological testing while retaining the assurance that compounds with novel shapes or properties have not been overlooked.

A shape-based clustering method has recently been shown to be useful for reducing the size of a database while retaining the geometrical diversity (42). When applied to test databases, timing studies indicate that the method is applicable to large datasets. One advantage of the distance method is that, for the same compound, conformational differences will be evident. Although this method does not require precomputing the conformations and exacerbates the problem of already large overhead of data storage. The connectivity method clusters together molecules of similar connectivity, without dependence on the conformation of the molecules, and therefore may not be an adequate representation of shape in systems capable of extending over a wide range of distances with a high degree of flexibility.

The application of the shape similarity approach, using atom triplets as descriptors, has also been used for rapid quantitative shape matching between two molecules or molecules and a template (93). The overall similarity is measured as a scoring factor between atom triplets.

Conclusions - There has been considerable progress in access and retrieval of 3D structural database information to aid in finding and designing new biologically active compounds. Three-dimensional database retrieval methods can already take into account the conformational flexibility of molecules. Efficient similarity searching methods are now aiming toward direct application to structure-activity and molecular design, wether at the level of small organic molecules or protein structures. In the current docking methods limited conformational flexibility, when addressed, is restricted to the ligand molecules. In general, docking or de novo strategies depend on the accuracy of the binding site. While the success of many of the methods depends on the starting

localization of the ligand with respect to the binding site, efficient sampling of the binding protein within reasonable computational limits ne ds improvement. Future developments will no doubt address these limitations.

STRUCTURAL DATABASES AND SIMILARITY SEARCHES

PROGRAMS	AFFILIATION	Reference
MOLPAT	Merck	10
QUEST3D/VISTA	Cambridge Crystallographic Database Centre	13
MACCS3D	Molecular Design Limited	13,14
CHEMDBS3D	Chemical Design Ltd.	15
UNITY	Tripos Associates, Inc.	16
CAST	Chemical Abstracts Service	17
FCD	Fine Chemical Directory	6
CONCORD	Tripos Associates, Inc.	18,19
COBRA/WIZARD	NA	20
ALADDIN	Daylight Chemical Information Systems	21
EMBED	NA	22
ASP	Oxford Molecular Limited	28-31

STRATEGIES IN 3D SIMILARITY SEARCHING

PROGRAMS	APPLICATION - VALIDATION	Reference
ASP	steroids	32
MEPCONF	monoamine oxidase B substrates	33
SPERM	dihydrofolate reductase	38
	DNA binding molecules	39
SUPER	leukotriene D4 receptor antagonists	45
RECEPS/AUTOFIT	dihydrofolate reductase ligands	46
YAK	carbonic anhydrase, thermolysin	47
DISCO	dopaminergic and benzodiazepine agonists	48
KEY/LOCK/LOCKSMITH	hemoglobin allosteric modifiers	49
CAVITY	HIV-1 protease	50
POCKET	adipocyte lipid binding protein, malate dehydrogenase	51
GEPOL	alcohol dehydrogenase, ribosomal protein	53
DOCK	papain, carbonic anhydrase	58
	alpha-chymotrypsin	59
DOCK2	trypsin, chymotrypsin, subtilisin	60
	thymidylate synthase	61
YING	dihydrofolate reductase, p-hydroxybenzoate hydroxylate, catabolite gene activator protein, phosphoglycerate kinase, ribonuclease T1	65
DISDOCK	serine protease complexes	66
AUTODOCK	immunoglobulin McPC 603, chymotrypsin, lysozyme, aconitase	67
CAVEAT	tendamistat, somatostatin analogs	75
FOUNDATION	thermolysin inhibitor	76
GROW	aspartyl proteases	77
LUDI	HIV protease, dihydrofolate reductase	78
DELEGATE(BUILDER)	HIV-1 protease	79
CLIX	L226Q influenza-virus hemagglutinin mutant	80
GRID	phospholipase-A2, dihydrofolate reductase, insulin	81
GenStar	FK-506 binding protein, HIV protease, carbonic anhydrase	82
LEGEND/LORE	dihydrofolate reductase	83
PROTEP	ubiquitin, thioredoxin, heat shock protein	91

References

1. "Concepts and Applications of Molecular Similarity", Eds., M.A. Johnson, G.M. Maggiora, John Wiley & Sons, Inc., New York, (1990).
2. I.D. Kuntz, Science, 257, 1078 (1992).
3. R. P. Sheridan, A. III. Rusinko, R. Nilakantan, R. Venkataraghan, Proc. Natl. Acad. Sci., USA, 86, 8165 (1989).
4. Y.C. Martin,M.G. Bures, P. Willett, in Rev. Computational Chemistry, Eds., K.B. Lipkowitz, D.B. Boyd, VCH: New York, N.Y., 213 (1990)

5. Y. C.Martin, J. Med. Ch m., 35, 2145 (1992)
6 O.F. Guner, D.W. Hughes, L.M. Dumont, J. Chem. Inf. Comput. Sci., 31, 408 (1991)
7. K.S. Haraki, R.P. Sh ridan, R. Venkataraghavan, D.A. Dunn, R. McCulloch, Tetrahedron Comput. Methodol., 3(6C), 565 (1990).
8. C.A. Pepperrell, P. Will tt, J. Comput.-Aided Mol. Des., 5, 455 (1991).
9. T.R. Hagadone, J. Chem. Inf. Comput. Sci., 32, 515 (1992)
10. P. Gund, W.T. Wipke, R. Langridge, Comput. Chem. Res., Educ., Technol., 3, 5 (1974).
11. C.A. Pepperrell, A.R. Poirrette, P. Willett, and R. Taylor, Pestic. Sci., 33, 97 (1991).
12. P.J. Artymiuk, P.A. Bath, H.M. Grindley, C.A. Pepperrell, A.R. Poirrette, D.W. Rice, D.A. Thorner, D.J. Wild, P. Willett, J. Chem. Inf. Comput. Sci., 32, 617 (1992).
13. F.H. Allen, J.E. Davies, J.J. Galloy, O. Johnson, O. Kennard, C.F. Macrea, E.M. Mitchell, G.F. Mitchell, J.M. Smith, D.G. Watson, J. Chem. Inf. Comput. Sci., 31, 187 (1991).
14. F.H. Allen and O. Kennard, Chem. Design Auto. News,8,1 (1993).
15. N.W. Murral and E.K. Davies, J. Chem. Inf. Comp. Sci., 30, 312 (1990).
16. UNITY, Tripos Associates, Inc., 1699 S. Hanley Road, Suite 303, St. Louis, Mo., 63144
17. W. Fisanick, K.P. Cross, A. Rusinko III, Tetrahedron Computer Methodology, 3,(6C), 635 (1990).
18. R.S. Pearlman, Chem. Des. Automation News, 2, 1 (1987); CONCORD, available from Tripos Associates Inc., St. Louis MO.
19. A. Rusinko III, R.P. Sheridan, N. Ramaswamy, K.S. Haraki, N. Bauman, R. Venkarataghavan,J. Chem. Inf. Comp. Sci., 29, 251 (1989).
20. A. R. Leach, Pestic. Sci., 33, 87 (1991).
21. J.H. Van Drie, D. Weininger, and Y.C. Martin, J. Comput.-Aided Mol. Des., 3, 225 (1989).
22. G.M. Crippen, T.F. Havel, in "Chemometrics Research Studies Series", D. Bawden, Ed.; Research Studies Press:Letchwath (Wiley); New York (1988).
23. D.E. Clark, P. Willet and P.W. Kenny, J. Mol. Graphics, 10, 194 (1992).
24. O.F. Guner, D.R. Henry, T.E. Moock, R.S. Pearlman, Tetrahedron Computer Methodology, 3,(6C) 557 (1990).
25. P. Willett, J. Chemom., 6, 289 (1992).
26. R. Carbo, L. Leyda, M.,Arnau, Int. J. Quantum Chem., 17, 1185 (1980).
27. R. Carbo, L. Domingo, Int. J. Quantum Chem., 32, 517 (1987).
28. ASP, Automated Similarity Package, Oxford Molecular Ltd, The Magdalen Centre, Oxford Science Park, Sandford on Thames, Oxford OX4 4GA, United Kingdom.
29. E.E. Hodgkin, W.G. Richards, Int. J. Quant. Chem.: Quant. Biol. Symp., 14, 105 (1987).
30. A.C. Good, J. Mol. Graphics, 10, 144 (1992).
31. A.C. Good and W.G. Richards, J. Chem. Inf. Comput. Sci., 33, 112 (1993).
32. A.C. Good, S.S. So, W.G. Richards, J.Med.Chem., 36, 433-438 (1993).
33. F. Sanz, F. Manaut, J.A. Sanchez, and E. Lozoya, J. Mol. Struct. (Theochem.), 230, 437 (1991).
34. P.-L. Chau, P.M. Dean, J. Mol. Graphics, 5, 97 (1987).
35. P.M. Dean, P.-L. Chau, J. Mol. Graphics, 5, 152 (1987).
36. P.M. Dean, P. Callow, P.-L. Chau, J. Mol. Graphics, 6, 28 (1988).
37. P. Bladon, J. Mol. Graphics, 7, 130 (1989).
38. V.J. van Geerestein, N.C. Perry, P.D.J., Grootenhuis, C.A.G. Haasnot, Tetrahedron Comput. Methodol., 3(6C), 595 (1990).
39. P.D.J. Grootenhuis, V.J. van Geerestein, C.A.G. Haasnoot, M. Karplus, Bull. Soc. Chim. Belg., 101, 661 (1992).
40. C. Amovilli, R. McWeeny, J. Mol. Struct. (Theochem.), 227, 1 (1991).
41. G. Buchbauer, S. Winiwarter, P. Wolschann, J. Comp.-Aided Mol. Design, 6, 583 (1992)
42. G.W. Bemis and I.D. Kuntz, J. Comput.-Aided Mol. Des., 6, 607 (1992).
43. P.G. Mezey, J. Math. Chem., 7, 39 (1991).
44. P.G. Mezey, J.Chem.Inf.Comp.Sci., 32, 650 (1992).
45. R.B. Hermann and D.K. Herron, J. Comput.-Aided Mol. Des., 5, 511 (1991).
46. Y. Kato, A. Inoue. M. Yamada, N. Tomioka, and A. Itai, J. Comput.-Aided Mol. Des., 6, 475 (1992).
46b. C.A. Del Carpio, Y. Takahashi, S.-I. Sasaki, J. Mol. Graphics, 11, 23 (1993).
47. A. Vedani, J. Rec. Res., 13, 163 (1993).
48. Y.C. Martin, M.G. Bures, E.A. Danaher, J. DeLazzer, I. Lico, and P.A. Pavlik, J. Comput.-Aided Mol. Des., 7, 83 (1993).
49. G.E. Kellogg and D.J. Abraham, J. Mol. Graphics, 10, 212 (1992).
50. C.M.W., Ho, G.R. Marshall, J. Comput.-Aided Mol. Des., 4, 337 (1990).
51. D. G. Levitt, L.J. Banaszak, J. Mol. Graphics, 10, 229 (1992).
52. J. S. D laney, J. Mol. Graphics, 10, 174 (1992).
53. E. Silla, F. Villar, O. Nilsson, J.L. Pascual-Ahuir, O. Tapia, J. Mol. Graphics, 8, 168 (1990).
54. A.S. Smellie, G.M. Crippen, W.G. Richards, J. Chem. Inf. Comput. Sci., 31, 386 (1991).
55. F. Kuhl, G.M. Crippen, D. Friesen, J. Comput. Chem., 5, 24 (1984).
56. N. Kasinos, G.A. Lilley, N. Subbarao, I. Haneef, Prot in Eng., 5, 69 (1992).
57. I.D. Kuntz, J.M. Blaney, S.J. Oatley, R. Langridge and T.E. Ferrin, J. Mol. Biol. 161, 269 (1982)

58. R. DesJarlais, R.P Sheridan, G.L. Seibel, J.S. Dixon, I.D. Kuntz and R. V nkataraghavan, J. Med. Ch m., 31, 722 (1988).
59. K.D. Stewart, J.A. B ntley, M. Cory, Tetrahedron Comp. Methodol., 3(6C), 713 (1990).
60. B.K. Sh ichet, I.D. Kuntz, J. Mol. Bi l., 221, 327 (1991).
61. B.K. Shoichet, R.M. Stroud, D.V. Santi, I.D. Kuntz, K.M. Perry, Sci nce, 259, 1445 (1993).
62. A. Badel, J.P. Mornon, S. Hazout, J. Mol. Graphics, 10, 205 (1992).
63. D.J. Bacon and J. Moult, J. Mol. Bio., 225, 849 (1992).
64. A.D. McLachlan, J. Mol. Biol., 128, 49 (1979).
65. S.L. Chan, P.-L. Chau, and J.M. Goodman, J. Comput.-Aided Mol. Des., 6, 461 (1992).
66. S. Y. Yue, Protein Eng., 4, 177 (1990).
67. D.S. Goodsell and A.J. Olson, Proteins: Struct., Funct., Genet., 8, 195 (1990).
68. T.N. Hart and R.J. Read, Proteins: Struct., Funct., Genet., 13, 206 (1992).
69. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth and A.H. Teller, J. Chem Phys., 21, 1087 (1953).
70. A. Calflish, P. Niederer and M. Anliker, Proteins: Struct., Funct., Genet., 13, 223 (1992).
71. J. Cherfils, J. Janin, Curr. Opinion in Struct. Biol., 3, 265 (1993).
72. B.K. Shoichet, D.L. Bodian and I.D. Kuntz, J. Comput. Chem., 13, 380 (1992).
73. J. Cherfils, S. Duquerroy, J.Janin, Proteins: Structure, Function, and Genetics, 11, 271 (1993).
74. F. Jiang, S.H. Kim, J. Mol. Biol. 219, 79 (1991).
75. P.A. Bartlett, G.T. Shea, J.J. Telfer, S. Waterman, in "Molecular Recognition: Chemical and Biological Problems", S.M. Roberts, Ed., Royal Society of Chemistry: London, 78, 182 (1989).
76. C.M.W., Ho, G.R. Marshall, J. Comput.-Aided-Mol. Des., 7, 3 (1993).
77. J.B. Moon and W.J. Howe, Proteins: Struct., Funct., Genet., 11, 314 (1991).
78. H.-J. Bohm, J. Comput.-Aided Mol. Des., 6, 593 (1992).
79. R.A. Lewis, D.C. Roe, C. Huang, T.E. Ferrin, R. Landgridge, and I.D. Kuntz, J. Mol. Graphics, 10, 66 (1992).
80. M.C. Lawrence and P.C. Davis, Proteins: Struct., Funct., Genet., 12, 31 (1992).
81. P.J. Goodford, J. Med. Chem., 28, 849 (1988).
82. S. Rotstein and M. Murcko, J. Comput.-Aided Mol. Des., 7, 23 (1993).
83. Y. Nishibata, A. Itai, Tetrahedron, 47, 8985 (1991).
84. P.-L. Chau, P.M. Dean, J. Comput.-Aided Mol. Des., 6, 385 (1992).
85. P.-L. Chau, P.M. Dean, J. Comput.-Aided Mol. Des., 6, 397 (1992).
86. P.-L. Chau, P.M. Dean, J. Comput.-Aided Mol. Des., 6, 407 (1992).
87. R. A. Lewis, J. Mol. Graphics, 10, 131 (1992).
88. J. M. Thornton, S.P. Gardner, E. G. Hutchinson, in "Computer Modelling of Biomolecular Processes", J. M. Goodfellow and D. S. Moss, eds., Ellis Horwood Ltd., 10, 211 (1992).
89. A. Sali, J.P. Overington, M.S. Johnson, T.L. Blundell, *idid*, 11, 232 (1992).
90. J.S. Fetrow, S.H. Bryant, Biotechnology, 11, 479 (1993).
91. H.M. Grindley, P.J. Artymiuk, D.W. Rice, P. Willett, J. Mol. Biol., 229, 707 (1993)
92. M.F. Lynch, E.M. Rasmussen, P. Willett, G.A. Manson, G.A. Wilson, Biochem. Soc. Trans., 17, 856 (1989).
93. R. Nilakantan, N. Bauman, R. Venkataraghavan, J. Chem. Inf. Comput. Sci., 33, 79 (1993).